DATA COMMUNICATIONS MANAGEMENT

NEXT-GENERATION NETWORKS: PART 1

Keith Knightson

INSIDE

The Great Divide: The Shift from Circuit to Packet Technology; Useful Comparisons; Separation of Services from Transport; Naming and Addressing; Telephony Service; The Next-Generation Internet

INTRODUCTION

This article is the first in a three-part series describing next-generation networks (NGNs). This first article examines the significantly different properties of packet-based NGNs and legacy circuit-based networks. These differences will determine the new trends of usage and characteristics of NGNs. Part 2 will present a set of graphical paradigms to illustrate the dramatic effects that NGNs will have on the communications environment. Part 3 will identify the effects that these paradigm shifts will have on users, network service providers, and the communications industry in general.

THE GREAT DIVIDE: THE SHIFT FROM CIRCUIT TO PACKET TECHNOLOGY

The main difference between traditional telecommunications networks and next-generation networks is the shift from a circuit-switched infrastructure to a packet-switched infrastructure. A number of defining characteristics can be identified for circuit-switched technologies. Similarly, a number of different characteristics can be identified for packet-switched technologies. In addition, the new applications and deployment requirements make demands not originally envisaged when the "first"-generation packet networks were designed.

Thus, in terms of effects and impacts arising from NGN deployment, it is necessary to briefly compare the major differences between circuit-switched and

PAYOFF IDEA

A number of traditional fundamental assumptions are being challenged by the disruptive developments of NGN, which will obsolete a number of existing architectural foundations and corresponding inter-organizational relationships. Consideration needs to be given to the development of new frameworks to accommodate these developments. For example, particular areas identified include protocol layering, interworking, end-to-end services (especially QoS), control and management, and even policy.

Auerbach Publications

packet-switched infrastructures, together with the related new and future demands.

Because all services — voice, data, video, etc. — are being digitized, it clearly is possible to carry them all on a single data network. It has been assumed, and may seem obvious, that the Internet and the Internet Protocol (IP) should form such a single platform[1] for all forms of information transfer (i.e., voice, data, and video). Thus, another major difference is the shift away from multiple vertically integrated networks of the past (e.g., voice networks, data networks, and video networks) toward a single horizontally integrated network of the future.

Given the "universality" of the Internet and its associated infrastructure, which includes Domain Name Servers, addressing schemes, and address assignment, and applications such as e-mail, file transfer, the World Wide Web, etc., it seems reasonable to assume that IP-based systems will form the basis of NGNs.

It is hardly conceivable that any other "system" is going to replace all this in the immediate future. Yet, as pointed out by Braden et al.,[25] there are aspects of the Internet and related technology that leave much to be desired in terms of serving the emerging environment and its new demands.

The Internet Protocol (IP) was not originally designed with real-time applications in mind. Some European operators are considering what they call separate "managed-IP-networks" for providing voice services. In this case, separate IP-based networks will be specifically engineered and interconnected according to particular end-to-end and intermediate performance criteria, and other criteria appropriate for provision of voice services.

In the same way that the telephone network has evolved over so many years, including many "revolutionary" events, the Internet will have to evolve to meet the new demands. Thus, it would be wrong to conclude that the extant Internet is the NGN. A more appropriate conclusion would seem to be that the Internet will evolve into the NGN.

USEFUL COMPARISONS

The following sub-sections compare some of the key characteristics for circuit-switched networks and packet-switched networks. This is not intended to be a tutorial on circuit-switched and packet-switched technologies, but merely an abstraction of the relevant fundamental properties and characteristics. Some poetic license has been taken in the descriptions to keep the information as understandable as possible to an audience that may not be familiar with, or interested in, the fine details of all the technologies involved.

General Considerations

Circuit-Switched Networks. Simplistically, a circuit-switched technology allows the establishment of dedicated and constant bandwidth "pipes" from one user to another. The constancy is both in terms of end-to-end delay and throughput. The establishment of the pipe (a.k.a. call) normally involves a call setup procedure, a data transfer phase, and a call tear-down procedure. Call records are

maintained during lifetime of the call, recording time of call establishment, duration, time of tear-down, and other information regarding call-related features (bandwidth, facilities, etc.).

Packet-Switched Networks. Users of packet-switched networks are neither interconnected by dedicated bandwidth nor invariant delay pipes. Users must share the bandwidth of packet trunks (and even access links) and the packet processing (switching) resources. An overload of either of these resources can result in the loss (discard) of data or the delay of data. There is no way to control the rate of user input, and this, coupled with the effect of random traffic patterns, can lead to erratic values of end-to-end delay (sometimes known as jitter). This fact is of particular concern to certain applications, such as voice, that require short and reasonably constant delays to be viable.

Connection-Oriented versus Connectionless Packet Switching. The Internet is based on the Internet Protocol (IP), which is a *connectionless* protocol.[3] Each and every packet contains all the information necessary for it to be routed from source to destination. Thus, every packet contains both source and destination addresses and, perhaps, other information that might be used to further affect the routes to be taken. IP is a datagram protocol and, as such, does not require any state information to be set up or maintained within the packet switches themselves. State information to keep track of user–user sessions is confined to end-user systems. IP was also designed to operate as an overlay network, over any underlying transmission technology. It does not require or use any specific features of any specific underlying technology. There are two flavors of IP: version 4 (IPv4) and version 6 (IPv6). The address space (range) of IPv4 has become exhausted and IPv6 has been designed to accommodate a much larger address space (range). Additionally, some new quality-of-service (QoS) signaling capabilities have been added.

Transparency

Circuit-Switched Networks. Once established, a circuit-switched "pipe" can be used for the transfer of content under the control of the users. Within the constraints of the pipe, the user is free to transfer information in any networking protocol of his choice. From an architectural perspective, transparency begins at layer 1.

Packet-Switched Networks. In the case of an IP network, true transparency only begins above the IP layer. As will be discussed later, in complex cases, there may be many packet layers or sublayers below the IP layer, all of which require or constrain users to implement a "stack" of protocols common to both the users terminal and the respective peer network elements. The protocols for layers up to and including IP are common to the user and the service provider. This tight coupling, and protocol dependency, has major implications on both the users and service providers.

Interference

Circuit-Switched Networks. Once a call (pipe) has been established between two users, there is no interference with, or from, other calls established in the network. The use of this "pipe" is not affected by other pipes that may be established between another pair of users (and vice versa).

Packet-Switched Networks. The very nature of a packet network creates *interference* among customers. This is extremely difficult, if not impossible, to avoid. Packet bearers are shared *statistically* on a contention basis. However, non-average or other unpredictable activity can cause overload,[4] which results in packets being deliberately discarded. This discard may be random or selective (if some precedence scheme is in operation).

The impact of interference can only be mitigated, to a certain extent, by the introduction of adequate QoS mechanisms and *fairness* mechanisms.

Congestion

Circuit-Switched Networks. The only kind of congestion that can occur is the inability to establish a pipe in the first instance. Once established, the bandwidth of the pipe is constant, the delay is constant, and no congestion of traffic will be experienced.[5]

Packet-Switched Networks. Packet networks can become congested if the offered load is higher than that provisioned for. Instantaneous overloads can be "smoothed out" by adequate storage, but only at the expense of extending the transit delay time. Longer-term overloads will result in packet loss as indicated above, that is, the random or selective discard of packets.

For real-time applications, it is desirable that the Internet be engineered to provide high levels of availability similar to those currently encountered in traditional telecommunications networks. Certain applications are more affected by temporary congestion than others. The impact of congestion can only be mitigated by the introduction of adequate QoS mechanisms, and possibly differentiation between traffic belonging to different types of application.

Quality-of-Service (QoS)

Circuit-Switched Networks. QoS is not, in general, a significant issue with circuit-switched networks. Traditional telecommunications networks have very high values for parameters such as availability, reliability, time-to-repair, etc. Rarely is dial tone not obtainable. Most "busy" signals are due to the end user not being available.

The quality of circuits, in terms of throughput or delay, is rarely an issue. Once a circuit has been established, the end-to-end delay is almost[6] negligible, and in all cases is constant.

It should be noted that data loss due to congestion does not occur on a circuit-switched network.[7] Data can only be submitted at the (*a priori*) understood rate supported by the "pipe" and this rate is constantly available.

Packet-Switched Networks. In a connectionless packet-switched network, QoS is more problematic[8] (than in a circuit-switched network) due to the random or erratic sharing of a single resource.

With a circuit-switched network, it is relatively easy to provide a given constant bandwidth with little transit delay and invariant delay for the lifetime of a given connection. No equivalent property exists in a packet-switched network.

With connectionless packet switching, the original concepts assumed no relation between packets. Packet switches do not "know" which packets constitute a particular flow between two particular remote end users, and no information is recorded; that is, no state information is retained for future use.

Traditionally, the Internet has offered what is euphemistically called a "best-effort" service. This property stems from the military origins of the Internet. The intent was to route around failed nodes; try any or all possible routes to a given destination. It was designed to operate under adverse conditions, including very low QoS.

All traffic is treated equally well — or equally badly as the case may be — according to available resources. There are those who would claim that, in these days huge of bandwidths and virtually unlimited memories and processing resources, the best-effort service is still sufficient. They will claim that such things are merely commercial dimensioning issues and the market will decide which ISPs (Internet service providers) are suitable and which are not.

However, it is known that certain applications will require a certain bandwidth, and/or an acceptable transit delay, to be viable. It is generally accepted that to provide a QoS for voice similar to that available on the traditional PSTN (public switched telephone network),[9] an end-to-end transit delay of no more than 100 milliseconds is required. A major question, therefore, is how to accommodate applications that, to be viable, must be supplied with adequate bandwidth and acceptably low transit delay.

Best-Effort Service. Because all "pipes" are shared by interleaving packet from numerous different sources, any one user can potentially affect, and be affected by, all other users. So-called "bandwidth hogs" can dominate flows to the detriment of other users, and long packets can affect transit delay times for others because they prevent transmission of other packets while they themselves are being transmitted. Until recently, these features were not problems as such, but simply the characteristics of a packet network and quite adequate for non-time-critical applications such as e-mail or file transfer. The original intent of the "Internet" was simply to make a "best effort" to deliver the data, with delivery itself being regarded as more important than minor delays or temporary congestion.[10]

Better-Effort Service. More recently, of course, with the desire to migrate voice services and other real-time applications to packet networks, the characteristics described above became unacceptable.

Mechanisms have been introduced to "segregate" different types of traffic such that the resource requirements for different traffic types can be engineered. Diff-serv (differentiated service) is one example; it allows traffic to be directed to paths that have been engineered with specific QoS parameters in mind. In this case, however, the flows are at a *macro* level, and each path is still shared by other traffic of similar kind. Additionally, the word "differentiated" implies that differentiated accounting charging will have to be achieved. The typical examples quoted are "gold," "silver," and "bronze" services, or similar marketing phrases. The semantics of such services are unclear; the standards only speak of techniques for distinguishing between relative streams and not of the characteristics of such streams. Thus, the use of diff-serv by applications remains unclear. The situation is further complicated by the fact that such streams only apply between adjacent points and not to the entire end-to-end path.

The efficacy of this approach has yet to be proven. In the context of a real-time, delay-sensitive application (voice, for example), solutions are required to:

- Relate differentiated services to the specific QoS requirements of specific applications
- Ensure end-to-end coherence (e.g., is the silver service of ISP X offered to a sender exactly equivalent to the silver service of ISP Y offered to the recipient and taken into account by all in between?)
- Solve the accounting/charging problem

The ability of the Internet to carry high volumes of real-time traffic (i.e., to completely replace the existing PSTN) is still an open question, both from technical and organizational viewpoints. The introduction of NGNs into the enterprise or campus environment is definitely viable, however.

Mobility Services

Circuit-Switched Networks. There is a well-developed infrastructure for the provision of mobility services. Number portability[11] is also possible within certain geographical limits. It should be noted, however, that in the case of wireless services, the "air" portion of the path may be "packet-like" in nature, inasmuch as the path is shared among a number of users who could potentially cause mutual interference with respect to each other's traffic.

Packet-Switched Networks. Standards for mobility have been developed for use at the IP layer; that is, in relation to an IP address. Higher-layer mobility, at the personal identity level, will be covered by the telephony service overlaid on top of the IP network, using Session Initiation Protocol- (SIP) based techniques or H.323-based techniques and related directory/mobility application services.

Accounting

Circuit-Switched Networks. Accounting for usable bandwidth is relatively easy because it can be based on the "size" (i.e., the bandwidth) of the pipe and duration (i.e., the length of the call) of the establishment of the pipe. Because the bandwidth of the pipe is fixed, providers generally do not charge on the basis of "useful" data actually transmitted, although there may be time periods when the pipe is idle. It is technically difficult to have volume-based accounting because the pipe is assumed to be transparent and no distinction can be made by the provider between the bits representing no traffic and bits representing useful user data.

An important aspect of a circuit-switched network is the use/availability of call records, achieved from the "state-information"[12] inherent in the operation of circuit-switched networks. Information about the source and destination addresses, as well as other information, is maintained for the duration of the call. At the end of the call, such information can easily be transferred to a storage medium for any subsequent processing, as for example, for accounting purposes (or for law enforcement purposes).

Packet-Switched Networks. Generally speaking, no time or volume accounting has been applied to the packet transfer in connectionless packet networks,[13] and particularly not to individual flows (i.e., the packets related to a particular a single source or destination pair). With millions or thousands of millions of packets per second flowing through a packet switch, recording every packet for subsequent processing would be impractical. No "call"-based records exist in a connectionless packet-switched environment, and one of the credos of the Internet community is to avoid the need for state information anywhere in the network (except in the end-systems).

Because no "circuits" are readily identifiable, any attempt to monitor "conversations" or trace the source or destinations thereof would require processor-intensive correlation between a random set of individual packets in packet switches (i.e., routers). For this reason, many ISPs charge flat rates for the exchange of traffic based on the size of address space of the customer site and the size of the customer access pipe.[14]

One major problem with charging for packets is how to "account" for dropped packets that have already been "counted" for charging to the sender but not actually delivered. This makes it very unlikely that per-packet charging is viable. Even charging based on some assumption about the senders rate is suspect in the face of potential packet loss after transmission. Additionally, verification of the actual loss rate would be difficult for most users to achieve.

Emergency Services

Circuit-Switched Networks. Users of the PSTN are familiar and accustomed to the availability of special arrangements for access to police, fire, or ambulance services by dialing 911. If a dial tone is available, then access to these services is almost guaranteed due to the nature of circuit-switched technology. Special

arrangements and resources are provided by the network providers for this service.

Packet-Switched Networks. It is not clear how emergency services would be provided in a packet-switched environment. For example, the segregation of emergency traffic in an Ethernet access network presents a particular challenge. In core networks, some kind of preference scheme could probably be engineered. However, it is more difficult to provide such services in a packet network than it was in a circuit-switched network where the users can control the access circuit, and where special dial-in numbers can be used to provide downstream separation. Moreover, it is not clear that ISPs would voluntarily implement any kind of emergency service unless required to do so by force of regulation.[15]

These issues are beginning to be addressed in various standards organizations. The same situation and considerations arise here as for Lawful Interception (see immediately below).

Lawful Interception

Circuit-Switched Networks. Traditionally, there are means and mechanisms for law enforcement agencies to monitor particular access circuits and "live conversations," and provide numerical and geographical location information relative to communicating parties.

The use and availability of call records, achieved from the "state-information," are inherent in the operation of circuit-switched networks. Information about the source and destination addresses, as well as other information, is maintained for the duration of the call. At the end of the call, such information can easily be transferred to a storage medium for any subsequent processing (e.g., for accounting or for law enforcement purposes).

Packet-Switched Networks. Because no "circuits" are identifiable, any attempt to monitor conversations or trace source or destinations thereof require processor intensive correlation between a random set of individual packets to become a standard feature in packet switches (i.e., routers).[16]

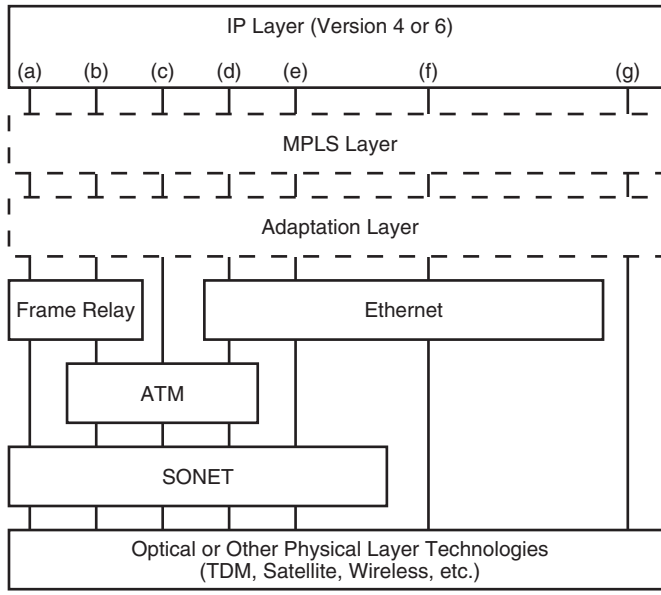
As above, it is not clear that ISPs would, in fact, voluntarily implement any kind of emergency service unless required to do so by force of regulation.

The Internet Engineering Task Force (IETF) has published an opinion that "the IP service-industry has no obligation to design, develop, or deploy IP protocols specifically to meet Lawful Interception requirements." However, these issues are beginning to be addressed in some standards organizations, notably within the European Telecommunications Standards Institute (ETSI). ETSI TR 101 944[17] provides a good discussion of the issues.

It should be noted that the separation between service and network considerably complicates matters. Both elements must be considered and work synergistically. The separation creates both technical and administrative problems. The use

of a dynamic or temporary IP address, now common practice, is also a problem for location tracking.

EXHIBIT 1 — Example of Protocol Stacks within an IP Network



Relationships with Other Network Technologies. There are several types of packet-switched networks. For example, all of the following are based on packet technology:

- X.25 Networks
- Frame Relay
- Local/metropolitan area networks (LANs/MANs); for example, Ethernet
- Multi-Protocol Label Switching (MPLS)

Additionally, there are network technologies, such as SONET or other optical systems, TDM, etc.

Any and all of the above can be utilized in an IP network. IP can operate over these technologies and provides top-level overlay network spanning such technologies to provide and act as the common, single, end-to-end protocol. Specific RFCs have been developed to specify the adaptation[18] required to operate IP over the various technologies.

The development of MPLS is interesting because it can be used to enhance the provision of QoS features for IP, by providing different connection-oriented flows underneath IP. ATM can be used in a similar manner, as can any flow that provides a "guaranteed" or some preferential treatment for applications that require it.

Exhibit 1 shows the general way in which various technologies can be integrated into an IP network. The resulting protocol stacks can be complex, as well as different in the different regions/portions of a given network. From the protocol stack perspective, these stacks can be applied to both the core and access segments of a given network. Each of the vertical lines a through g represent a specific combination of protocols arranged one on top of the other. The exhibit is not exhaustive, being simply for illustrative purposes. Other protocols and combinations are possible.

SEPARATION OF SERVICES FROM TRANSPORT

The shift from circuit-switched technology to packet-switched technology is a major aspect but not the only one. Another key aspect is the separation of service aspects from the transfer aspects. This is a key issue in achieving so-called "convergence." Traditionally, there were specific networks for specific services; for example, PSTN for voice; IP, X.25, Frame Relay for data, and cable networks for video. In their original form, these networks were "tailored" in some sense for the service to be carried. Convergence expresses a desire to carry all services on a single NGN. For this to be possible, it is clear that the network should be service transparent and service independent.

Consequently, it also means a separation between service and network. But what does this mean in real practical terms? It implies that the mechanisms for service provision will no longer need to be "embedded" or "bundled" with the transfer mechanisms. That is to say, different services could be provided by using different service platforms. Some service providers could specialize in some services and not in others. There could be a choice of different service providers for a given service.

This separation has profound consequences on traditional services and related aspects. This will be discussed in more detail in Part 3. Consider, for example, the provision of emergency services such as fire, police, ambulance, etc., and even lawful interception, These are generally bundled together with telephone service and the underlying network. If services are separated from the network, how will such services be provided, and by whom? It is clear that some IP network providers may not provide telephone services.

The phrase "separation of service from network" may be misleading and/or misunderstood. It could be taken to mean that services are not part of NGN (the network). The word "network" in the phrase NGN should be viewed as the umbrella term for both the "services network" and the "transport network." These are the two elements that are to be separated.[19]

NAMING AND ADDRESSING

Telephone numbers are a key element to the functioning of the telephone network. Their importance lies in three key attributes:

1. A universally agreed-upon and unambiguous numbering scheme
2. The allocation of such numbers to end users
3. Network mechanisms that allow any user to reach any other user on the telephone network.

It is clear that an NGN must use an addressing scheme with similar global properties. The format telephony numbering plan is hierarchical in nature, structured into regions, countries, areas, local, etc. The benefit of such a hierarchy is its scaling properties. For example, the allocation or re-allocation of local numbers does not affect the format of the country in which the number change is made. Thus, routing tables operating at the inter-country level are not affected by local changes. Furthermore, inter-country routers only need routing to hold the country values and nothing more. A flat (i.e., randomly allocated) unstructured address space would require every switch to hold all possible addresses, and any change anywhere would affect all routers in the network.

In multi-layer network architectures, addressing can occur at several layers. Usually, however, there is only one layer which is an end-to-end layer, and at which truly global addressing is used. In the other layers (if they exist), the addresses are usually only of local significance. With regard to the Internet, the IP address is the address that fulfills the three requirements.

LAN Media Access Control (MAC) addresses could also fulfill such requirements. However, systems based exclusively on using MAC addresses do not scale well, and LANs are not usually global in geographical scope (although they could be in theory), nor is there a global directory service for them. LAN switches or bridges rely on "learning" and storing local LAN addresses from remote LAN stations, from the global set of LAN addresses allocated by IEEE. These addresses are not hierarchical in nature and thus require that every address has to be "learned." LANs can play an important role as collector networks, in limited geographical areas where only a relatively small number of addresses is involved. This technique can also alleviate the need for an equivalent number of IPv4 addresses, which are in short supply.

The extant Internet is neither hierarchical nor geographical in terms of address allocation and routing. Rather, its individual organizational parts are integrated on the basis of domain-oriented address and name allocations.

TELEPHONY SERVICE

Most of the previous sections dealt primarily with "connectivity" services rather than application services. This is not surprising, given that the use of packet networks and the digitization of all services (e.g., voice, data, video) intrinsically

decouples the carriage (transport) network from the application service platforms and the associated (digital) data[20] that would be used to provide such services.

However, because the telephony service was previously inextricably bound up with its carriage network (i.e., the PSTN), it is important to cover how such services will be provided if the PSTN is to be replaced by a packet network. This subject will be discussed in more detail in Parts 2 and 3 because it is one of the fundamental impact areas. For the time being, it is sufficient to say that the telephony service[21] will be separated from the transport network.

The subject of naming and addressing also comes up when considering provision of a telephone service over an IP network and when considering interworking between the PSTN-based telephone service and an IP-based telephone service. A solution termed "ENUM" has been formulated for "translation" of a number-based scheme into the (domain) name-based schemes generally used for application services in IP-based networks.

THE NEXT-GENERATION INTERNET

It is important to realize, and understand, that turning the Internet into the NGN is not entirely straightforward.

The following extract,[2] coming from renowned experts inside the Internet Community with long involvement with its development, is instructive in this regard:

> The design of today's Internet technology was guided by an Internet architecture that was developed in the 1970s, under the Internet research program of the Defense Advanced Research Projects Agency (DARPA) of the U.S. Department of Defense. Current reality and changing requirements are eating away at the viability of the original Internet architecture. Much of the coherence of the original architecture is being lost in a patchwork of technical embellishments, each intended to satisfy a particular new requirement.

It is clear that, as an NGN, the Internet would have to deal with requirements and environments for which it was not originally designed. Some recent "band-aid" developments have fixed some problems but created others, distorting coherence of the network, and perhaps even to the extent of preventing other requirements being met.

For example, a new architecture might be designed to create greater functionality, generality, adaptability, and/or robustness in the future Internet. On the other hand, without a new, long-term technical road map, the Internet is likely to become decreasingly effective, often failing to meet the demands placed on it by applications.

As a result of the study proposed by Braden et al.,[25] under DARPA funding, for USC/ISI,[22] MIT/LCS,[23] and ICSI[24] are collaborating on a research project to reconsider the Internet architecture in the light of present realities and future requirements.

The original peer-to-peer and end-to-end concepts in the original "architecture" have already been violated by the use of Network Address Translation

(NAT) functions and Security Firewalls as extensions, which themselves interfere with the use of IPSec (IP Security) extensions.

> As a result, current IETF protocol engineering is in a conceptual and technical muddle that will continue to lead to increasingly complex and costly engineering as well as loss of functionality. The only way to avoid this degeneration will be to restore coherence to the architecture, and that will happen only as the result of a deliberate effort to create a new architecture for the future. [See also Braden et al.[25]]

The following new (NGN) requirements that would influence a new Internet architecture are identified as follows:

- Mobility
- Policy-driven auto-configuration
- Highly time-variable resources
- QoS offering, management, and related accounting requirements
- Performance management
- QoS offering and management

Other problems include:

- The universal adoption and transition to IPv6
- Effect of MPLS
- A muddle of numerous tunneling schemes

A detailed discussion of these issues is beyond the scope of this article; however, further information can be found in Braden et al.[25] and IETF RFC 2775.[25]

This brief outline should suffice to indicate that we cannot simply say that NGN is already here because we have the Internet.

CONCLUSION

A number of traditional fundamental assumptions are being challenged by the disruptive developments of NGN, which will obsolete a number of existing architectural foundations and corresponding inter-organizational relationships. Consideration needs to be given to the development of new frameworks to accommodate these developments. For example, particular areas identified include protocol layering, interworking, end-to-end services (especially QoS), control and management, and even policy.

Significant changes are underway in terms of new and emerging technologies. Coupled with the effects of competition, these changes have surprising and profound technical and societal effects. These factors combine to represent tectonic shifts in the world of provision of telecommunications services as implemented by NGNs. These effects will be further identified and discussed in Parts 2 and 3.

Notes

1. While this might appear to be stating the obvious, it is by no means a foregone conclusion. While there is no doubt about the use and efficacy of the Internet for data applications, the wholesale replacement of the existing public switched telephone network is not entirely straightforward. Some people believe that other solutions such as ATM (Asynchronous Mode Transfer) are better suited for telephony services by virtue of their Quality of Service features. MPLS (Multi-Protocol Label Switching) has also been cited as a possible candidate to carry voice services without an intervening IP layer.
2. Brian Carpenter, RFC 2775, Internet Transparency, February 2000.
3. For simplicity, this article uses the terms "connection-oriented" (CO) and "connectionless" (CL) to distinguish between the two classical modes of operation. CO packet protocols and network technologies (such as X.25) are, in general, those in which closely coupled call setup, data transfer, and call tear-down phases are present. In CL protocols/networks, these distinct per user–user conversation phases do not occur. In CL networks, packets, sometimes termed "datagrams," always contain both the source and destination addresses.
4. The whole point of packet switching is the sharing of bearer based on statistical assumptions, and thus the need for less resources than a circuit-switched network. Provisioning of a packet network for maximum traffic would require roughly the same resources as the equivalent number of circuit-switched network calls.
5. Lack of dial tone could be regarded as a kind of congestion, as could fast busy (encountered when all trunks are busy).
6. Except in the rare instance where a satellite hop is included in the end-to-end path.
7. Noise can cause packet loss due to corruption of the data. However, because the same issue occurs for packet bearers, this is not singled out as a differentiating characteristic.
8. QoS has always been a difficult issue for packet networks. Even for connection-oriented packet networks such as X.25, "effective" QoS mechanisms were never implemented despite the definition of service features and parameters. Frame Relay and ATM have fared better with their "guaranteed" bit rate services.
9. Traditionally, the QoS for PSTN has been tightly defined in terms of overall end-to-end performance requirements, with an apportionment of parameters to various network segments (i.e., the international portion, the national portion, the local portion, etc.). The parameters were specified in relation to a "hypothetical reference connection." The Internet is not structured in the same fashion, being a more loosely coupled set of IP networks. Controlling the end-to-end performance, or even establishing suitable parameters, may be much more difficult in such an environment.
10. One of the original applications for an IP network was for the U.S. Department of Defense, where survivability was a key requirement, even under adverse circumstances.
11. Number portability is regarded as a mobility service for fixed access terminals and users, who change location or service provider from time to time. However, there are initiatives to integrate fixed and wireless mobility systems with a global range.
12. State-information is the term used to represent the information generated and required to be maintained during the lifetime of the instance of a connection, regarding the phases of the connection and related parameters. Examples would include establishment in progress, connection accepted/completed, resources allocated, data transfer phase entered, tear down requested, etc. Separate state-information is generated and maintained for each call (instance). See reference Marvin Minsky, "Computation — Finite and Infinite machines" for theory of state machines.
13. Connection-oriented networks such as X.25 can provide volume-based accounting based on call records.
14. When PSTN dial-up access is used, accounting is often applied, based on the holding time of the PSTN connection to the network access server (NAS).
15. To date, the United States has declared that the Internet is not to be the subject of regulation (of any kind).
16. Or, alternatively a Point of Interconnection (POI) for a separate dedicated processor.
17. Telecommunications Security; Lawful Interception (LI); Issues on IP interception, ETSI TR 101 944.
18. Adaptation normally consists of mapping functions; but in some cases, some type of "shim" protocol layer may be required.
19. The same confusion arises when we speak of the "Internet." Do we mean just the IP network, or the IP network plus its application services (WWW, FTP, VoIP, etc.)?
20. This "data" is sometimes termed "content."

21. There are various terms used for voice service over IP networks, such as IP telephony, Voice-over-IP (VoIP), etc. Traditionally, the characteristics of "the telephone service" adopted by most countries/operators are the ones that have been defined by the ITU-T Study Group 2. In this article, we use the term "telephone service" in this sense.
22. University of Southern California, Information Sciences Institute Computer Networks Division.
23. Massachusetts Institute of Technology, Laboratory for Computer Science.
24. International Computer Science Institute.
25. Robert Braden, David Clark, Scott Shenker, and John Wroclawski, "Developing a Next-Generation Internet Architecture," July 15, 2000, http://www.isi.edu/newarch.

Further Reading

American Civil Liberties Union, "Technological Analysis of Open Access and Cable Television Systems," December 2001, http://www.internetctc.com.

S. Blake, RFC 2475, "An Architecture for Differentiated Services," December 1998.

ITU-T Recommendation X.200 (aka ISO/IEC 7498-1), "Open Systems Interconnection — Basic Reference Model."

Bob Braden, Architectural Principles of the Internet, IPAM Tutorial, March 12, 2002, http://www.isi.edu/newarch.

Keith G. Knightson has been involved in data communications for more than 25 years, and has worked for British Telecom and Nortel Networks. He has operated a consulting company since 1995, concentrating on network architectures and next-generation networks. He can be contacted at kgk@rogers.com.